CLEARANCE REQUEST FOR PUBLIC RELEASE OF DEPARTMENT OF DEFENSE INFORMATION

(See Instructions on back.)

(This form is to be used in requesting review and clearance of DoD information proposed for public release in accordance with DoDD 5230.9.)

TO: Director, Freedom of Information & Security Review, Rm. 2C757, Pentagon

1. DOCUMENT DESCRIPTION

a. TYPE Presentation for Posting to Website	b. TITLE Mathematical Models for Cost Estimation and Analysis
c. PAGE COUNT 48	d. SUBJECT AREA Cost Analysis

2. AUTHOR/SPEAKER

a. NAME *(Last, First, Middle Initial)* Lee, David A.	b. RANK CIV	c. TITLE Senior Research Fellow
d. OFFICE NA		e. AGENCY LMI

3. PRESENTATION/PUBLICATION DATA *(Date, Place, Event)*

Request permission to post the presentation to the DoD Cost Analyst Symposium website at http://www.dodcas.osd.mil. The purpose is to make symposium presentations available to interested parties. The symposium was held on February 14-18, 2005 at the Williamsburg Marriott in Williamsburg, VA.

4. POINT OF CONTACT

a. NAME *(Last, First, Middle Initial)* Angers, Jeffrey P.	b. TELEPHONE NO. *(Include Area Code)* (703) 692-8045

5. PRIOR COORDINATION

a. NAME *(Last, First, Middle Initial)*	b. OFFICE/AGENCY	c. TELEPHONE NO. *(Include Area Code)*

CLEARED
For Open Publication

APR 26 2005 5

Office of Freedom of Information
and Security Review
Department of Defense

6. REMARKS

7. RECOMMENDATION OF SUBMITTING OFFICE/AGENCY

a. THE ATTACHED MATERIAL HAS DEPARTMENT/OFFICE/AGENCY APPROVAL FOR PUBLIC RELEASE *(qualifications, if any, are indicated in Remarks section)* AND CLEARANCE FOR OPEN PUBLICATION IS RECOMMENDED UNDER PROVISIONS OF DODD 5320.9. I AM AUTHORIZED TO MAKE THIS RECOMMENDATION FOR RELEASE ON BEHALF OF:

Chairman, Cost Analysis Improvement Group

b. CLEARANCE IS REQUESTED BY 20050431 *(YYYYMMDD)*.

c. NAME *(Last, First, Middle Initial)* Vogel, Russell A.	d. TITLE CAIG Executive Secretary
e. OFFICE OSD PA&E	f. AGENCY DoD
g. SIGNATURE Russell A. Vogel	h. DATE SIGNED *(YYYYMMDD)* 20050405

DD FORM 1910, MAR 1998 (EG) PREVIOUS EDITION MAY BE USED. WHS/DIOR, Mar 98

05-S-1274

LMI
GOVERNMENT CONSULTING
THE OPPORTUNITY TO MAKE A DIFFERENCE HAS NEVER BEEN GREATER

Mathematical Methods for Cost Estimating and Analysis

David Lee

38th Annual DoD Cost Analysis Symposium

February 2005

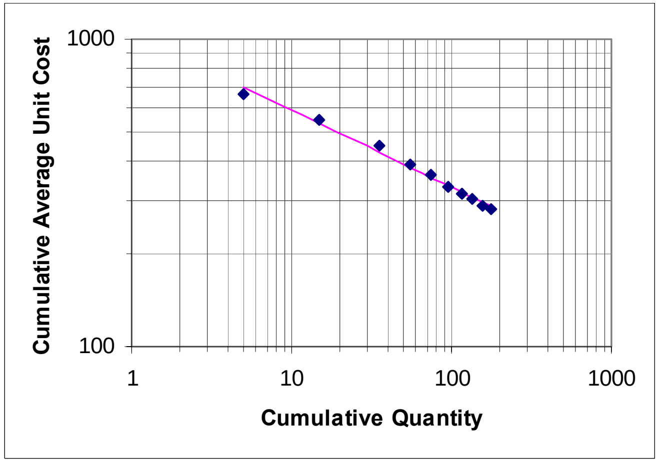
Two General Ideas

- When making and using mathematical models for cost estimating and analysis, one should:
  - Base the models on economic or physical principles, OR
  - Base the models on clearly described empirical evidence
  - Develop and apply the models with careful mathematics and statistics
- Benefits:
  - Maximal useful output
  - Straightforward explanations of the work
  - Ability to use discrepancies to improve the models
- Cost estimating and analysis generally belongs to the discipline of system identification, and methods of this discipline are useful to cost analysts

Three examples taken from "The Cost Analyst's Companion, illustrate these ideas

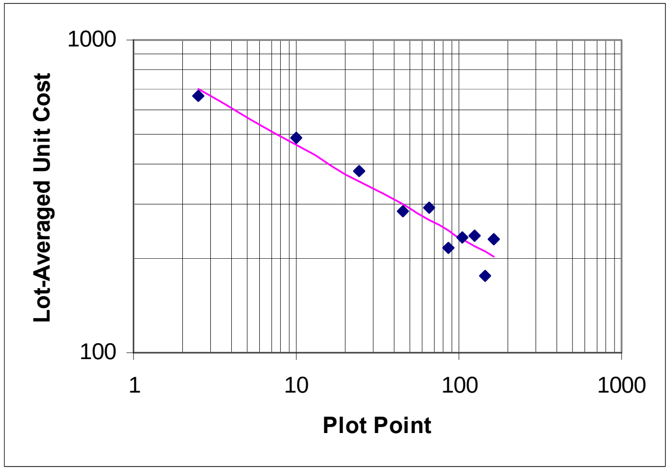
Analysis of Crawford Cost Progress Curves

- Wright and Crawford models are empirical



$$CAUC(Q) = \hat{T}_1 Q^{\hat{b}}$$

Wright



$$UC(j) = T_1 j^{b}$$

Crawford

Having chosen the Crawford model, use correct numerical values.

Efficient treatment of Crawford lot costs

When treating Crawford lots, use accurate numerical values of

$$A(L,U,b) \equiv \sum_{j=L}^{U} j^{b}$$

They're easy to generate as visual basic routines in MSExcel, or with Pascal or C/C++ routines. With them, one can treat parameter identification problems directly, without clumsy iteration and without using inaccurate approximations.

A(L, U, b), related to the Riemann zeta function $\varsigma(s) = \sum_{1}^{\infty} j^{-s}, \mathrm{Re}(s) > 1$**, is a special function of particular interest to cost professionals.**

Example VB routine

```
Function zze(l, u As Long, b As Double) As Double

Dim bb(1 To 10) As Double
Dim c, s, t, uu, ll, tu, tl, h, r, kk, p, q, uc, lc, u2, l2 As Double
Dim i As Integer

bb(1) = 0.16666667: bb(2) = -0.033333333: bb(3) = 0.023809523: bb(4) = -
0.0333333333: bb(5) = 0.075757576
bb(6) = -0.25311355: bb(7) = 1.16666667: bb(8) = -7.0921569: bb(9) = 54.971178:
bb(10) = -529.12424
c = 1 + b
uu = u + 0.5
ll = l - 0.5
uc = uu ^ c
lc = ll ^ c
u2 = uu ^ (-2): l2 = ll ^ (-2)
p = uc / c: q = lc / c
s = p - q
h = 2: r = 1: i = 0: kk = b + 2: ll = b + 3
tl = 1
Do
  h = 0.25 * h: i = i + 1: r = 2 * i * (2 * i - 1) * r
  kk = kk - 2: ll = ll - 2
  p = kk * ll * p * u2
  q = kk * ll * q * l2
  t = (1 - h) * bb(i) * (p - q) / r
  tl = Abs(t / s)
  s = s - t
Loop While ((tl > 0.00000001) And (i <= 9))

zze = s

End Function
```

User never sees this, of course; just inserts zze(L, U, b) in a MSExcel cell.

Typical Crawford Parameter Identification

Given a set of M lot costs, and the lots' start and end units, $\{c_i, L_i, U_i\}_1^M$ **,**

find T_1 and b of the "best fit" Crawford curve. Treat directly with

$$\min_{T_1, b} \sum_1^M (c_i - T_1 A(L_i, U_i, b))^2$$

or other minimization method. No need to use approximations like

$$A(L,U,b) \approx \frac{1}{1+b}\left[(U+0.5)^{1+b} - (L-0.5)^{1+b}\right]$$

and no need to use linear regression of ln(lot-averaged cost) on ln(plot point), with iteration on b.

LMI GOVERNMENT CONSULTING

Crawford Plot Points

Plot point $\overline{x}$ of lot with first unit L and last unit U is defined by

$$(U - L + 1)T_1\overline{x}^b = T_1\sum_{L}^{U} j^b = T_1A(L, U, b)$$

so

$$\overline{x} = \left[\frac{A(L, U, b)}{U - L + 1}\right]^{\frac{1}{b}}$$

LMI GOVERNMENT CONSULTING

What About a Rational Cost Progress Model?

- Cost progress comes from
  - Production workers' learning their tasks
  - Re-design of product for lower-cost production
    - Discrete elements embodied in integrated circuits
    - Re-designed structural members for cheaper production
  - Improved production facility
    - Better jigs, fixtures
    - Better layout
  - Lower-cost suppliers
    - Better make-buy decisions

All but the first of these require investments, and time, to make them happen.

When is it in the manufacturer's interest to make the investments?

Ingredients:

- Model of variation of unit cost with investment
- Model of demand schedule seen by manufacturer

Variation of Unit Cost with Investment

Let unit cost C vary with investment I as $C = f(I) = C^* + \Delta e^{-\alpha I}$

Derivation: $\Delta C = -\alpha\, \Delta I\,(C - C^*)$

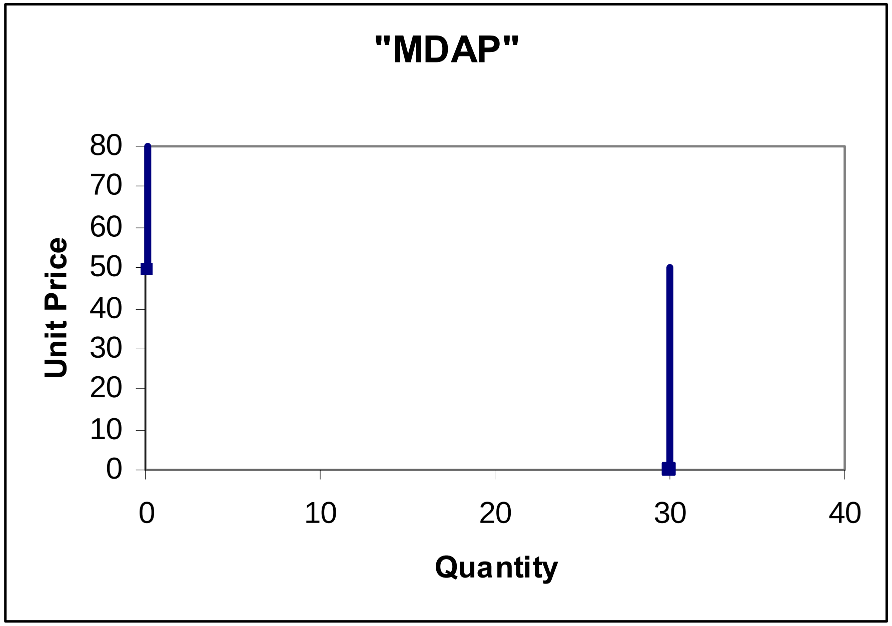
$$\frac{d(C - C^*)}{C - C^*} = -\alpha\, dI$$

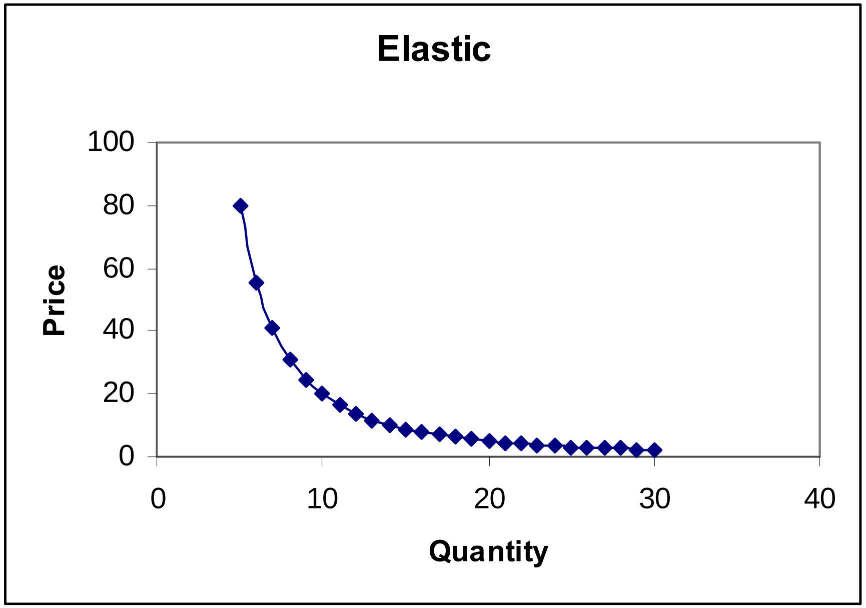
$$C(0) = C_0$$

$$C = C^* + (C_0 - C^*)e^{-\alpha I}$$

This f(I) builds in diminishing returns, and a minimum unit cost

Demand Schedules





"MDAP" and elastic demand schedules lead to similar results. We'll explore "MDAP."

Manufacturer's Optimization Problem

Choose investment sequence I_1, I_2, …, I_{N-1} and price sequence p_1, p_2, …, p_N to solve

$$\max_{\{I_j\}_1^{N-1},\,\{p_j\}_1^{N}} \sum_{j=1}^{N} \left\{ Q(p_j)\left[p_j - \Delta e^{-\alpha I_{j-1}} - c^*\right] - (I_j - I_{j-1}) \right\} b^j$$

subject to

$$\delta_{max} \geq I_j - I_{j-1} \geq 0;\ I_0 = 0;\ I_N = I_{N-1}$$

That is, choose investment and price sequences to maximize net present value of profit, subject to some obviously necessary constraints, and one not-so-obvious constraint. The δ_{max} constraint reflects the fact that inventing and implementing improvements takes time, and investments may be constrained by capital rationing, internal to the firm if not external.

Solution for “MDAP” Demand Schedule

For this demand schedule, manufacturer obviously sets price at the “tipping” price p_0 and builds the externally determined quantity Q_0. The solution of the optimization problem is

$$
\begin{aligned}
I_j &= j\delta_{max}, j = 1,2,\ldots,j^* \\
I_j &= \frac{1}{\alpha}\ln\left(\alpha\Delta Q_0 \frac{1-b^{N-j^*}}{1-b}\right) - (j^*-1)\delta_{max}, j > j^*
\end{aligned}
$$

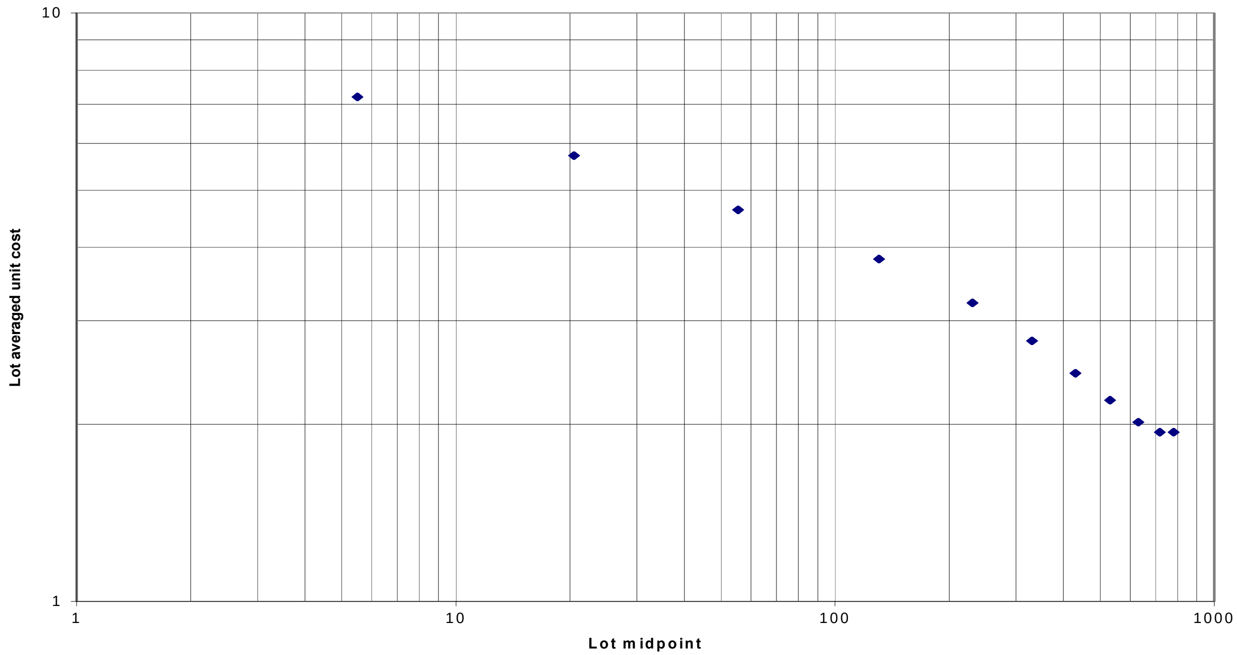
where j* is the smallest j such that

$$
\frac{1}{\alpha}\ln\left(\alpha\Delta Q_0 \frac{1-b^{N-j}}{1-b}\right) - j\delta_{max} \le \delta_{max}
$$

The unit cost sequence is

$$
\begin{aligned}
c_j &= \Delta e^{-\alpha(j-1)\delta_{max}} + c^*, 1 \le j \le j^* \\
c_j &= \Delta e^{-\alpha I_{j^*+1}} + c^*, j > j^* + 1
\end{aligned}
$$

Example

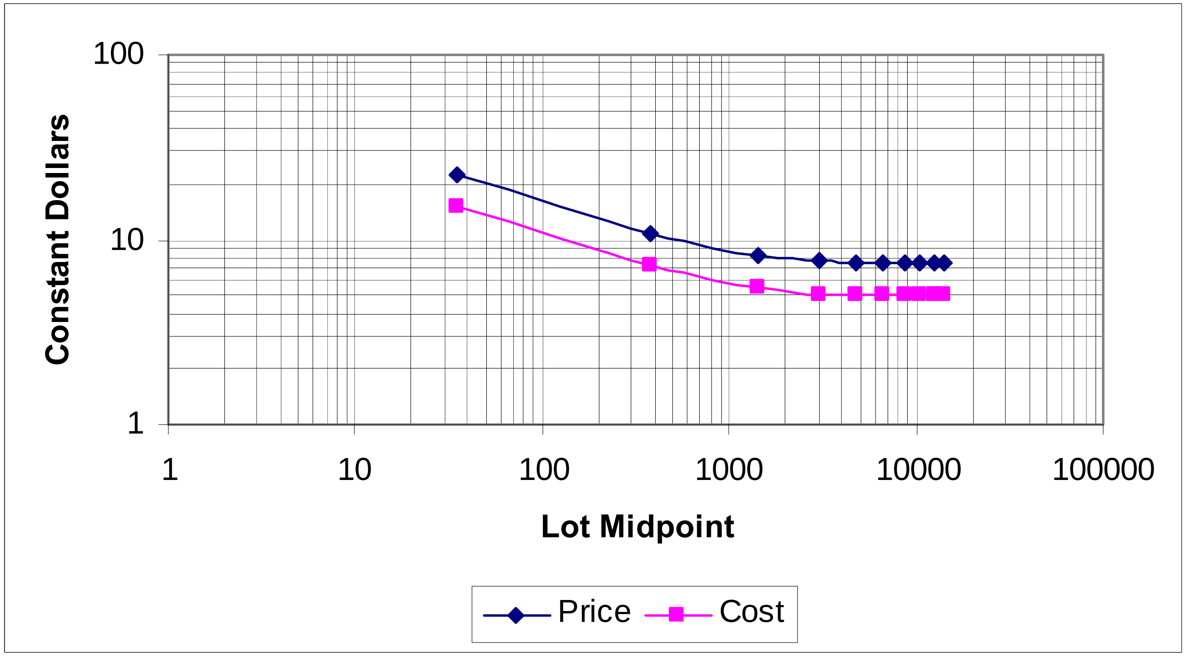


Typical “S” curve cost progress; cost progress eventually stops

A More Complex Example

For the "MDAP" example there is cost progress, but no price progress. More complex examples show similar investment patterns, and also exhibit price progress.

Optimal Price and Cost, Demand Schedule Elasticity ≡ 4



Curves Typically have three shape parameters

$$H \equiv \frac{\Delta}{C^*}$$ **("Headroom"; measures excess of initial cost over best cost)**

$$S \equiv \alpha Q_0 C^*$$ **("Sensitivity"; ratio of "good" lot cost to e-folding investment)**

$$L \equiv \alpha \delta_{max}$$ **("Limit"; ratio of maximum investment to e-folding investment)**

These, together with buy profile and the value of C*, determine the cost progress curve

Qualitative relations of parameters to product, production characteristics

$$\boxed{H \equiv \frac{\Delta}{C^*}}$$

Leads to larger H

- Hurried EMD; great time pressure for item
- Firm has little experience producing similar items

Leads to smaller H

- Substantially automated plant

H is large when production begins at unit cost well above best unit cost

Qualitative relations of parameters to product, production characteristics

$$\boxed{S \equiv \alpha \overline{N} C *}$$

Leads to larger S

- Flexible, relatively inexpensive tooling
- Many steps in production

Leads to smaller S

- Extensive, expensive specialized tooling
- Substantially automated facility

S is large when lot cost is large compared to e-folding investment

Qualitative relations of parameters to product, production characteristics

$$\boxed{L \equiv \alpha\delta_{max}}$$

Tends to larger L

- Product dominant in firm
- Competition or threat thereof
- Great confidence in total quantity

Tends to smaller L

- Sole-source procurement
- Uncertain future

Quantitative relations of parameters to product, production characteristics

- Three binary variables:
  - f_1: 1 => “complex” product
  - f_2: 1 => “automated” manufacturing
  - f_3: 1 => “competition” or threat thereof

Relating curve parameters to product and plant

Three translog functions:

$$H = H_0 \beta_1^{f_1} \beta_2^{f_2} \beta_3^{f_3}$$

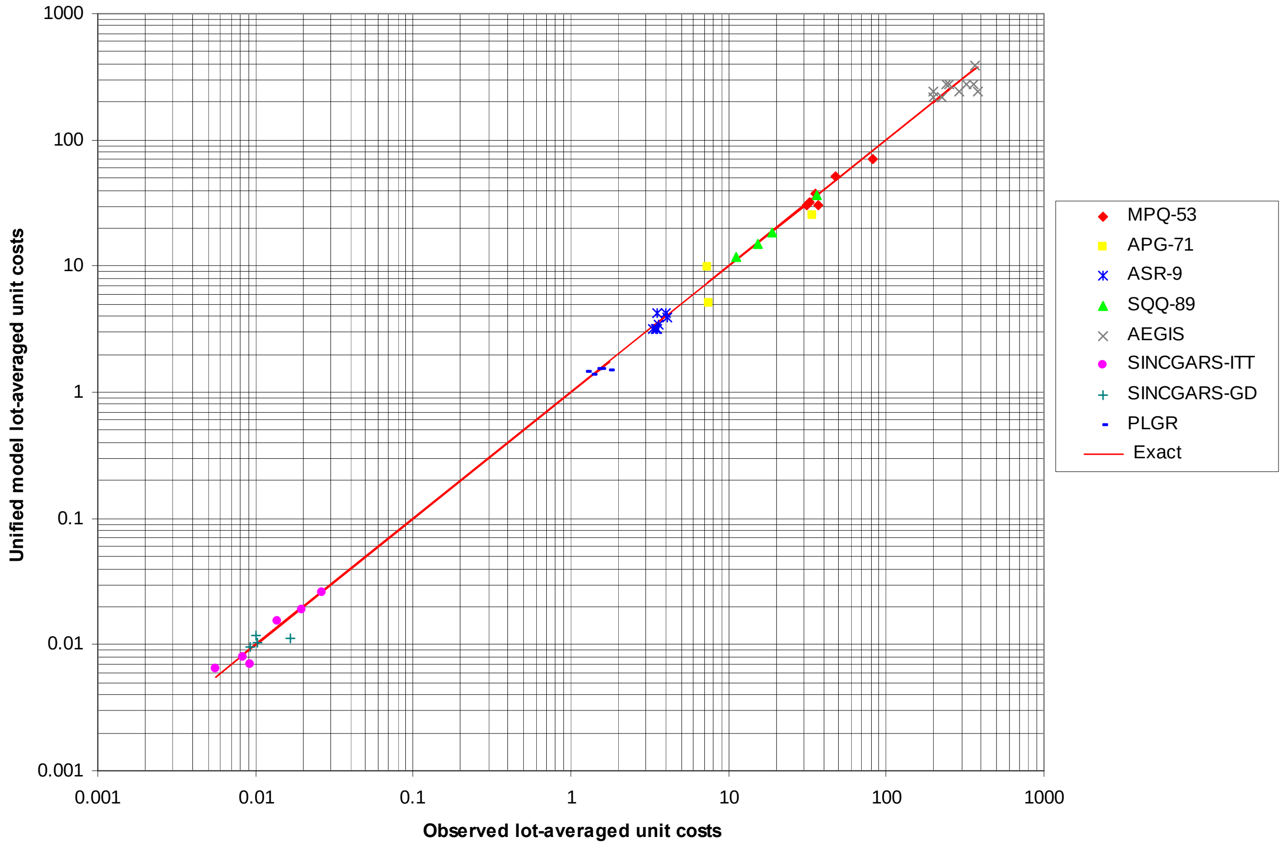
and similar translog functions for S and L.

Full disclosure: These are traditional functions, chosen arbitrarily!

(Developing rational model = research opportunity)

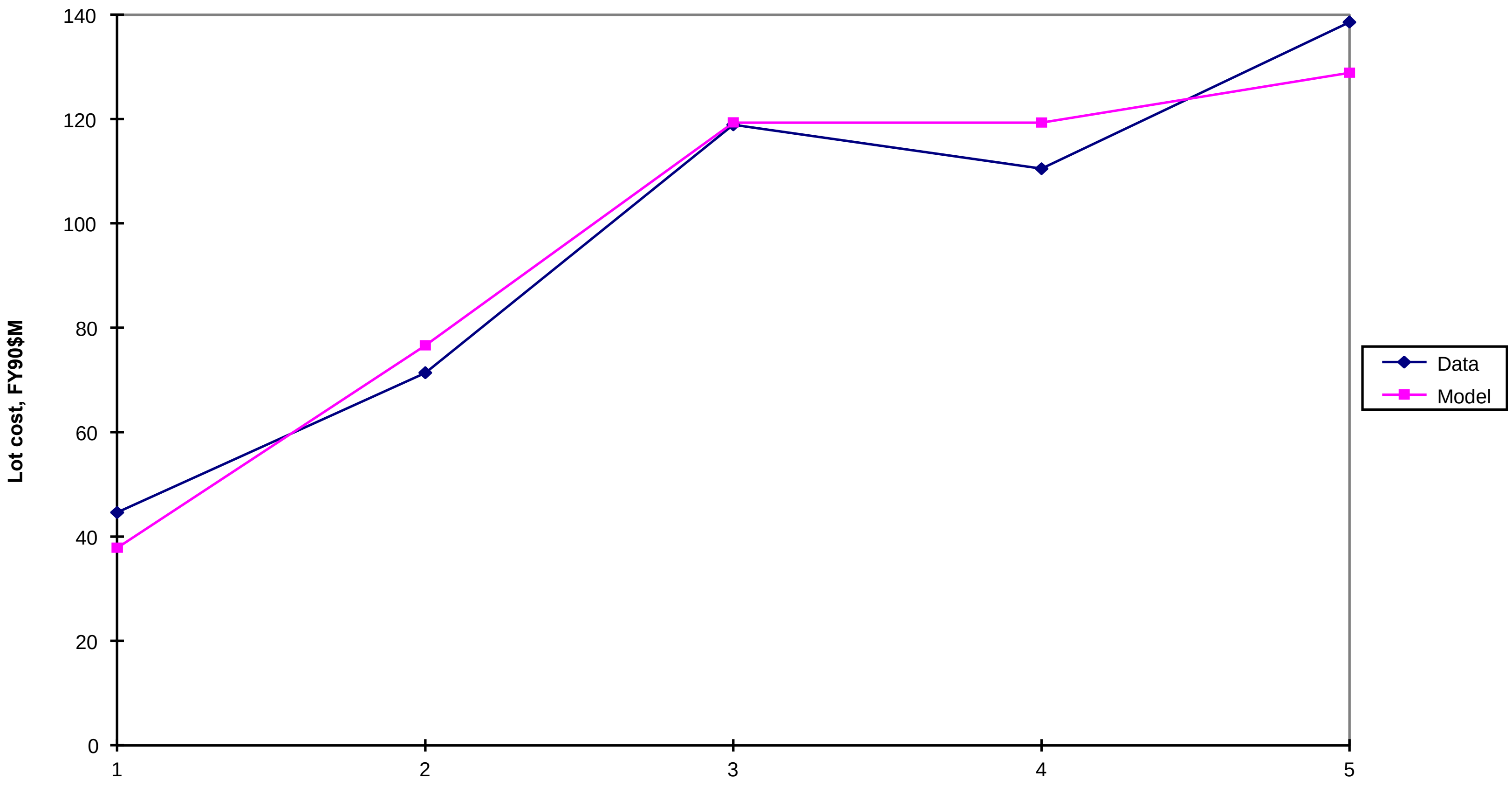
LMI GOVERNMENT CONSULTING

Results



1000
100
10
1
0.1
0.01
0.001
0.001
0.01
0.1
1
10
100
1000
Unified model lot-averaged unit costs
Observed lot-averaged unit costs
MPQ-53
APG-71
ASR-9
SQQ-89
AEGIS
SINCGARS-ITT
SINCGARS-GD
PLGR
Exact

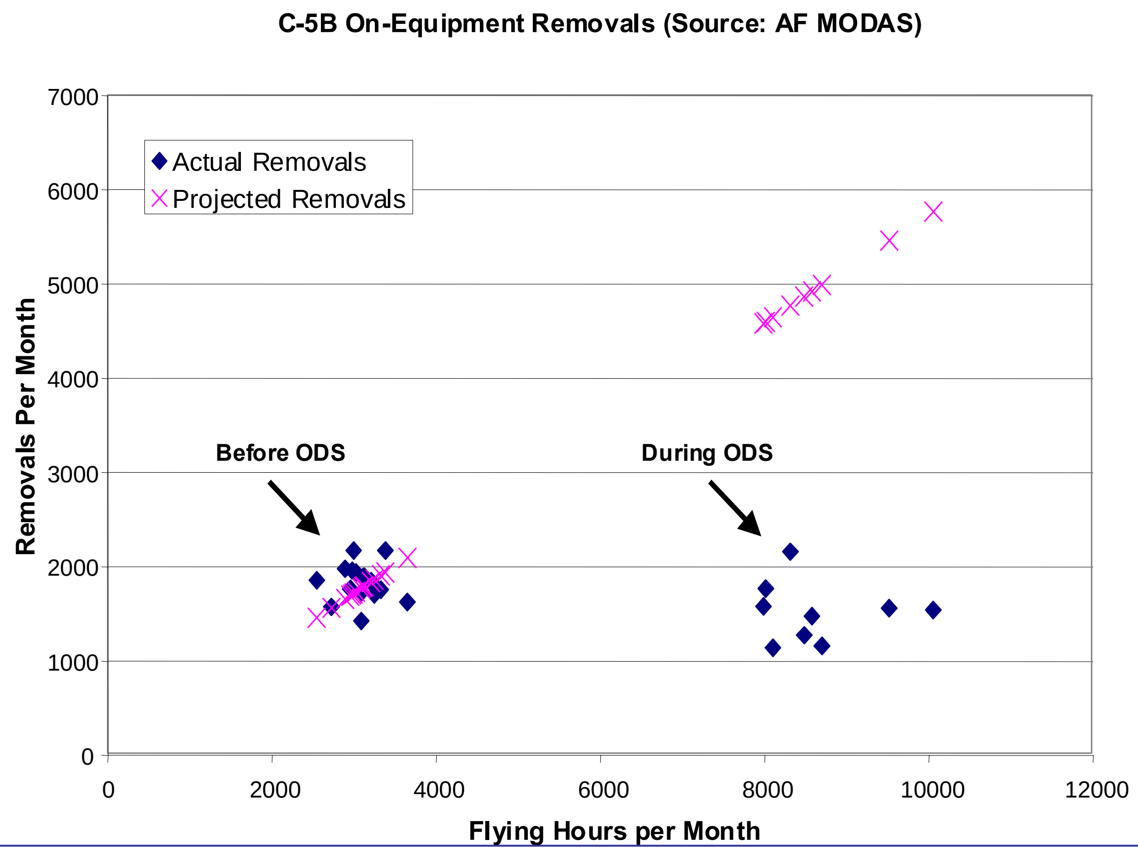
Application to a System Not Used in Calibration



Lessons learned

- When analyzing or forecasting cost progress, consider
  - Nature of the product
  - Nature of the manufacturing process
  - Business environment of the firm
- Use readily available processing power to apply rational (well, anyhow, partly rational) model

Aircraft Spares Requirements



J. Wallace: Modeling removals as simply proportional to flying hours over-predicted C-5B experience in Operation Desert Storm by more than 200%.

LMI GOVERNMENT CONSULTING

Statistical Models

Continuous Process: Exponential/Poisson

Physics: Each instant is equally likely to see a failure

Distribution of time between failures: $\lambda e^{-\lambda t}$ **(Exponential distribution)**

Probability of n failures in time T: $\frac{(\lambda T)^n}{n!} e^{-\lambda T}$ **(Poisson distribution)**

Expected failures: λT **Variance:** λT

Episodic Process: Binomial

Physics: Each event sees a failure, with probability P

Distribution of events between failures:

$$P(1-P)^j$$

Probability of n failures in M events: $\binom{M}{n} P^n (1-P)^{M-n}$

Expected failures: Mp Variance: MP(1-P)

A Physics-Based Model

- Each cold-start cycle in aircraft operations causes failures by a process having a binomial distribution with probability P_{cc}
- Each warm-start cycle causes failures by a process having a binomial distribution with probability P_{wc}
- Each daily cycle of temperature and humidity variations causes failures by a process having a binomial distribution with probability P_{gc}
- Flying induces failures by a process having the Poisson distribution with parameter λ_{FH}
- Each period has sufficiently many hours and cycles that approximating the discrete distributions by normal distributions with the same means and variances is acceptable
- Identifiable trends should be treated by standard means (e. g. Box-Jenkins-Reinsel shear transformation)

The Current Model

- In each observation period, aircraft experience Ncc cold cycles, Nwc warm cycles, and N_g diurnal cycles; they fly for t_f hours
- Mean and variance of the distribution of the number of failures:

$$\mu = N_g P_g + N_{cc} P_{cc} + N_{wc} P_{wc} + \lambda_f t_f$$

$$\sigma^2 = N_g P_g (1 - P_g) + N_{cc} P_{cc} (1 - P_{cc}) + N_{wc} P_{wc} (1 - P_{wc}) + \lambda_f t_f$$

Calibration

Given (Ncc, Nwc, Ng, t_f, removals) for periods 1, 2, …, M, estimate Pg, Pcc, Pwc, and λ_f: The present model is a 4-parameter model.

Note that variance of removals depends on Ncc, Nwc, Ng, t_f, so multilinear regression isn't a maximum-likelihood estimator. But it is simple enough to write down the likelihood of the observed sequence of removals, and calibrate the model by solving:

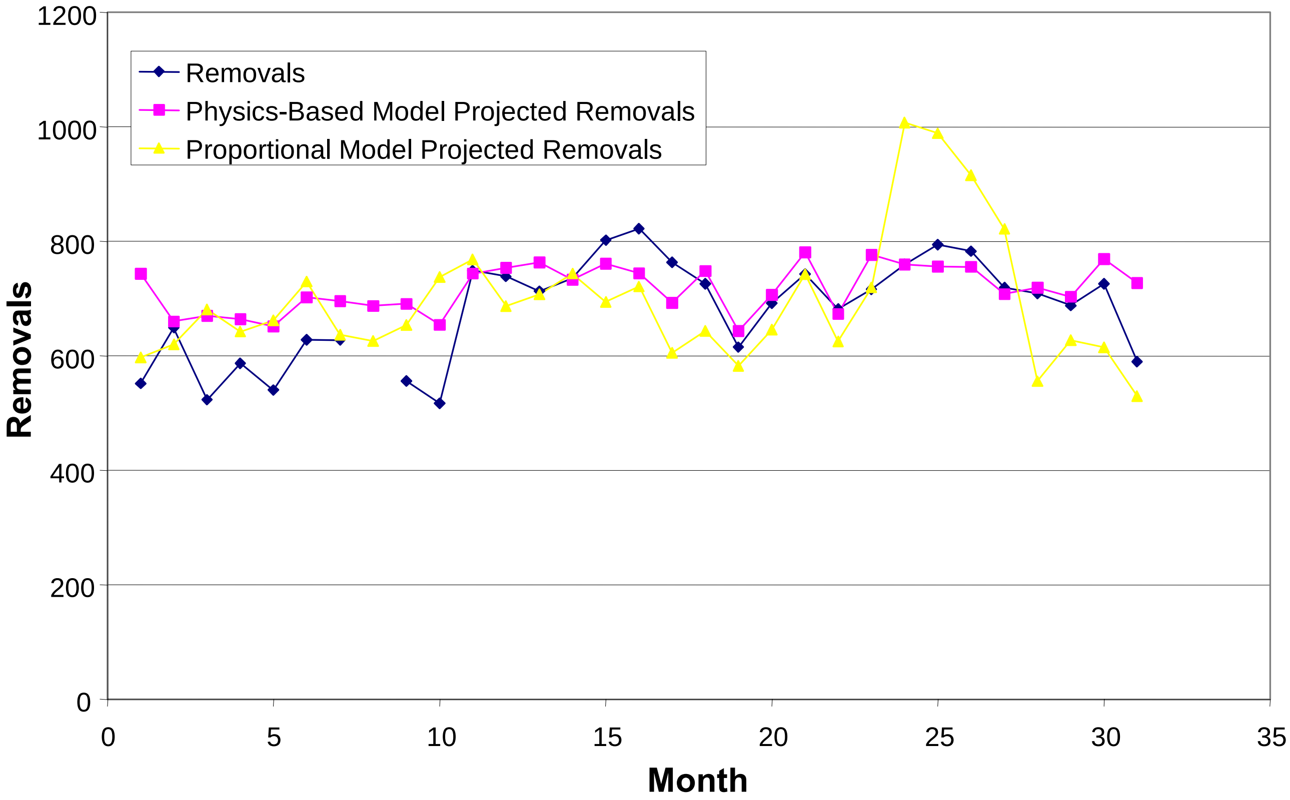
$$\min_{\bar{y}} \sum_{1}^{M} \left[\frac{(r_i - \mu(\bar{x}_i, \bar{y}))^2}{2\sigma^2(\bar{x}_i, \bar{y})} + \ln(\sigma(\bar{x}_i, \bar{y})) \right]$$

$$\bar{x}_i \equiv (Ncc_i, Nwc_i, Ng_i, t_{fi}); \quad \bar{y} \equiv (Pcc, Pwc, Pg, \lambda_f)$$

Solver in MSExcel may do the job. (It's worthwhile to check any optimizer's work.)

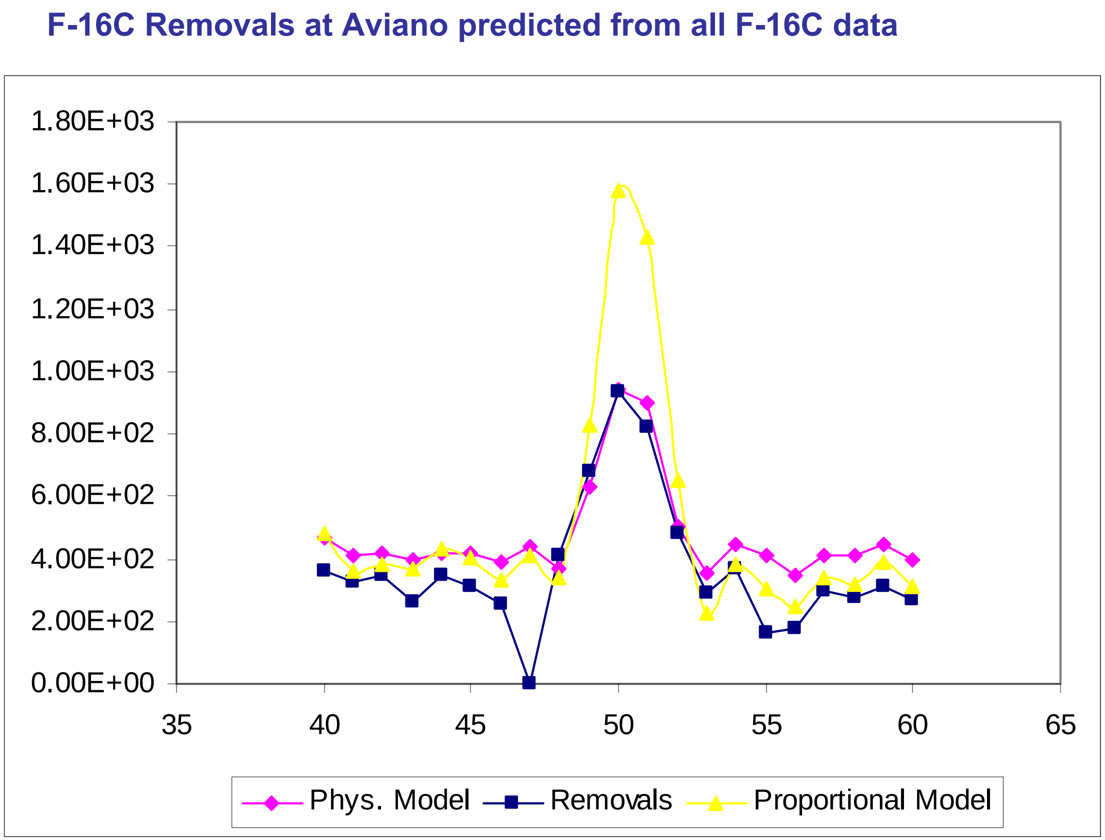
LMI
GOVERNMENT CONSULTING

Results: C-17 Fleet



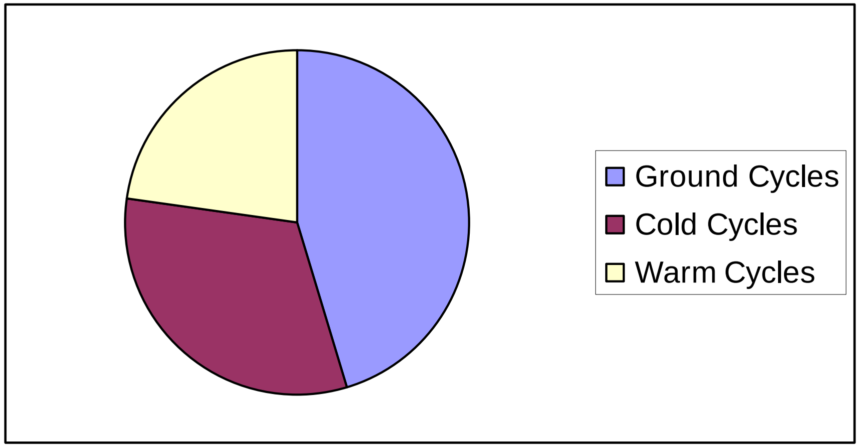
Calibration period: Months 16 – 25.   Explanatory variables: Ground cycles, warm cycles, and flying hours

F-16C Results

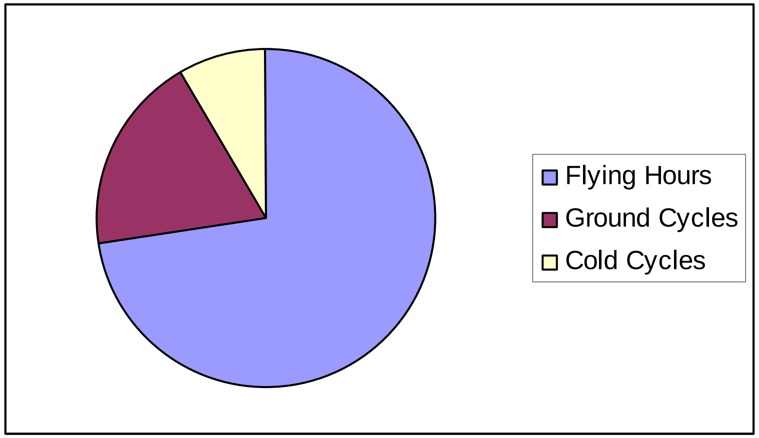


Physics-Based Models Give Helpful Information

C-5B

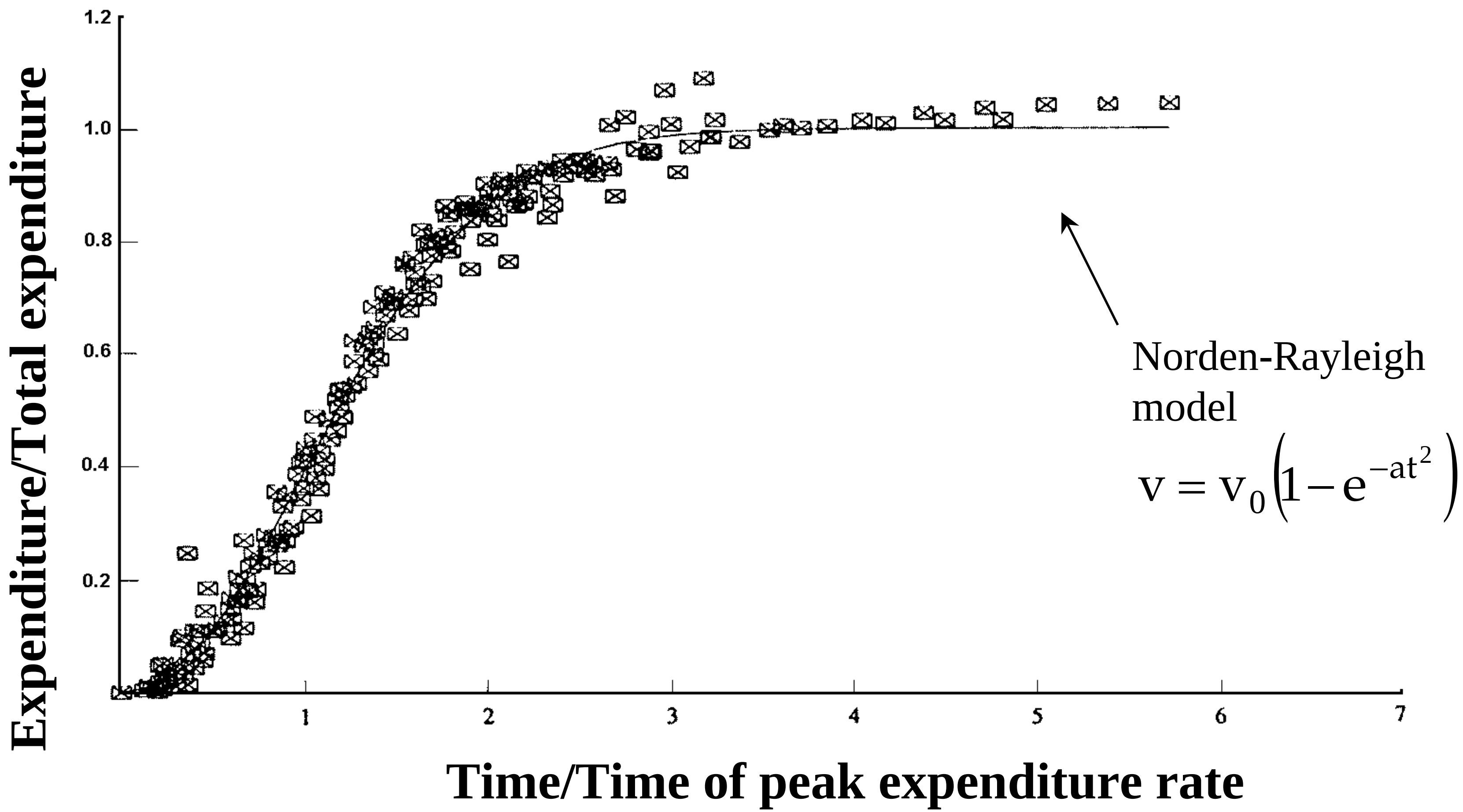


F-16C (Aviano)



Physics-based models show fractions of removals attributed to specific causes

The Norden-Rayleigh model collapses data from many DoD development programs onto one curve



Using the N-R model to estimate cost-to-go and time-to-go, given ACWP data

- Apply a parameter-identification method to estimate time-scale parameter a and cost-scale parameter d, with consistent estimates of dispersion (uncertainty). Many methods are available.
- Estimate completion time and total cost, with dispersion (uncertainty) estimates, from the a and d estimates.

One method: Apply MMAE

- System identification technique, developed for engineering work
- Applying it to EV analysis was Mark Gallagher's good idea
- Multiple Model Adaptive Estimation is a method for estimating parameters of dynamic systems, given time-history data.
- Uses set of Kalman filters, which require a parametric model for the time evolution of the system.

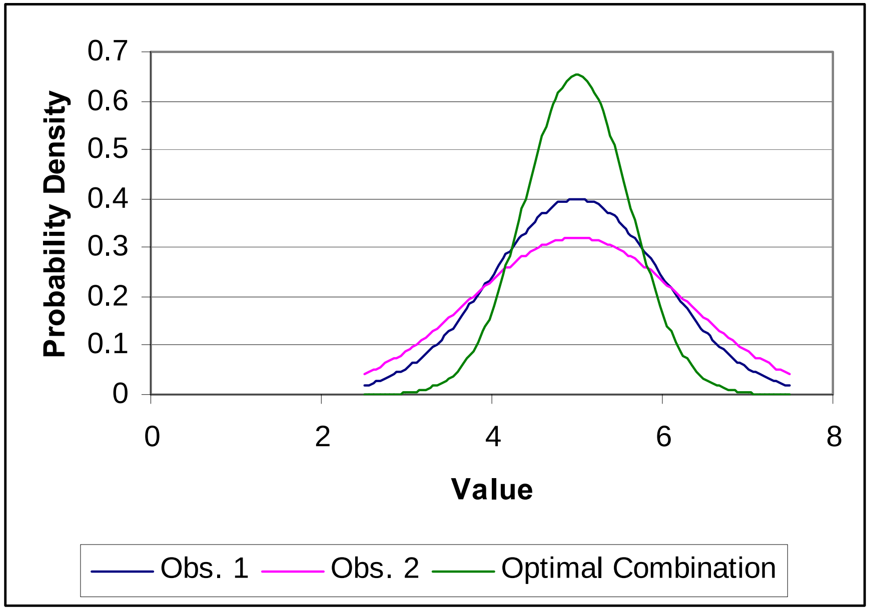
Gallager, M., and D. Lee, "Final-Cost Estimates for Research & Development Programs Conditioned on Realized Costs," Mil. Ops Rsch. 2, 1996, pp 51 - 65

Basis for Kalman Filter

$$z_1 \sim N(b,\sigma_1); z_2 \sim N(b,\sigma_2)$$

$$z_c = kz_1 + (1-k)z_2$$

$$k = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \Rightarrow \mathrm{var}(z_c) = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$



The variance of an optimal linear combination of two noisy observations is smaller than the variance of either observation

Kalman filter

- Given a system evolution model, Kalman filter estimates system state as a linear combination of the state predicted by the evolution model, and noisy observations of the state. For us, “state” is earned value v.

$$v(\mathrm{est}) = (1 - k)\, v(\mathrm{pred}) + k\, v(\mathrm{obs})$$

- Parameter k is called the gain of the filter

Maybeck, P., “Stochastic Models, Estimation and Control: Volume 1, Academic Press, New York, 1979

N-R time-evolution model

If $v = d[1 - \exp(-at^2)]$, then

$$\frac{dv}{dt} = 2ad\left(1-\frac{v}{d}\right)\sqrt{-\frac{1}{a}\ln\left(1-\frac{v}{d}\right)}$$

Evolution of earned value

If $v(t_0) = v_0$, then for $t > t_0$,

$$v = d\left[1 - e^{-a\left(t - t_0 + \sqrt{-\frac{1}{a}\ln\left(1 - \frac{v_0}{d}\right)}\right)^2}\right] \equiv V(t; a, d, t_0, v_0)$$

MMAE

- MMAE considers a bank of Kalman filters, each determined by three parameters (a, d, k), and determines probability that these are correct, given the ACWP data.

Maybeck, P., "Stochastic Models, Estimation, and Control: Volume 2 Academic Press, New York, 1982

Maybeck, P. S., and K. P. Hentz, "Investigation of Moving-Bank, Multiple Model Adaptive Algorithms," AIAA Journal of Guidance, Control, and Dynamics 10, 1987, pp. 771-101

Outputs from MMAE parameter identification

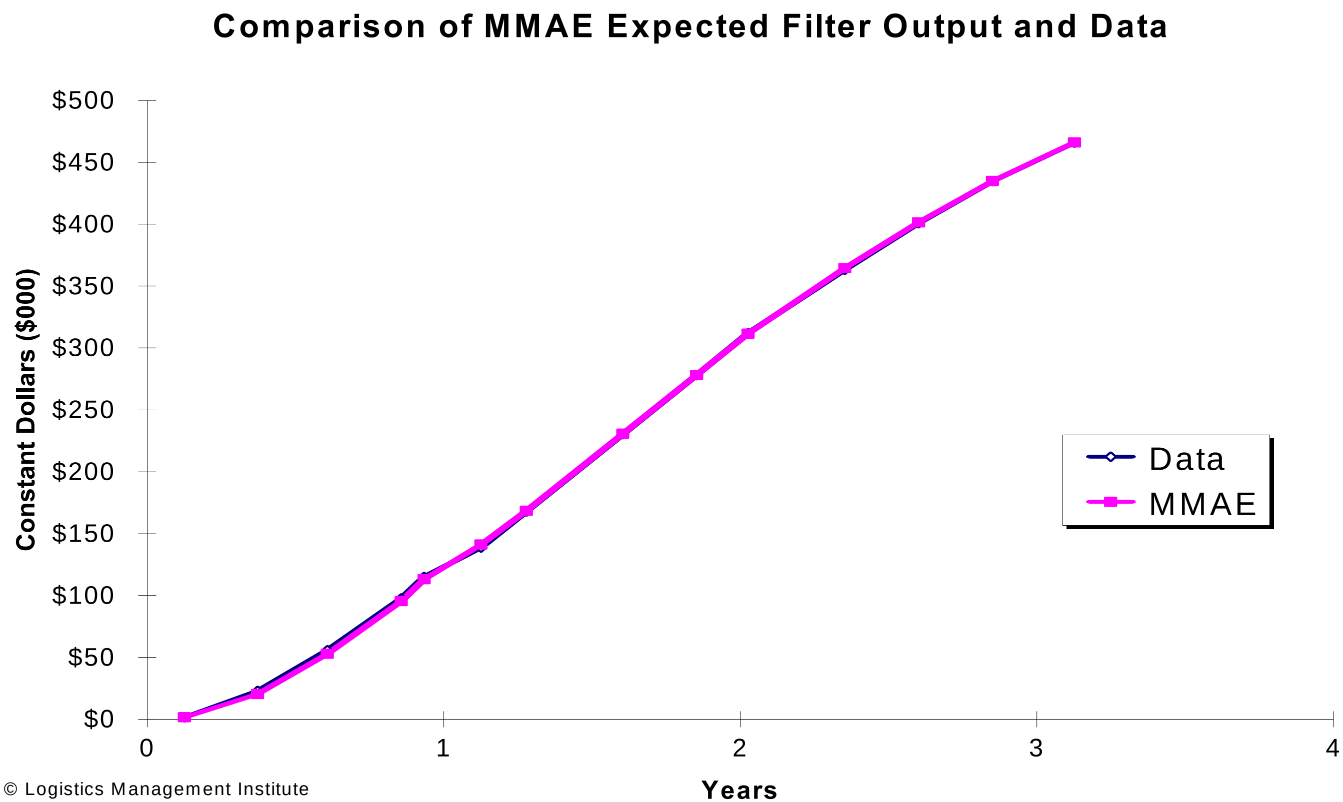
- Marginal distribution functions of total cost and total time, conditioned on the data
- Joint bivariate PDF of total cost and total time, conditioned on the data
- Can present costs either as $BY or as $TY

Full disclosure: the statistical analyses for MMAE are valid only for linear evolution equations. As others have done, we used the results anyway, for a non-linear equation. A remedy might be, to linearize the evolution equation about some representative solution.

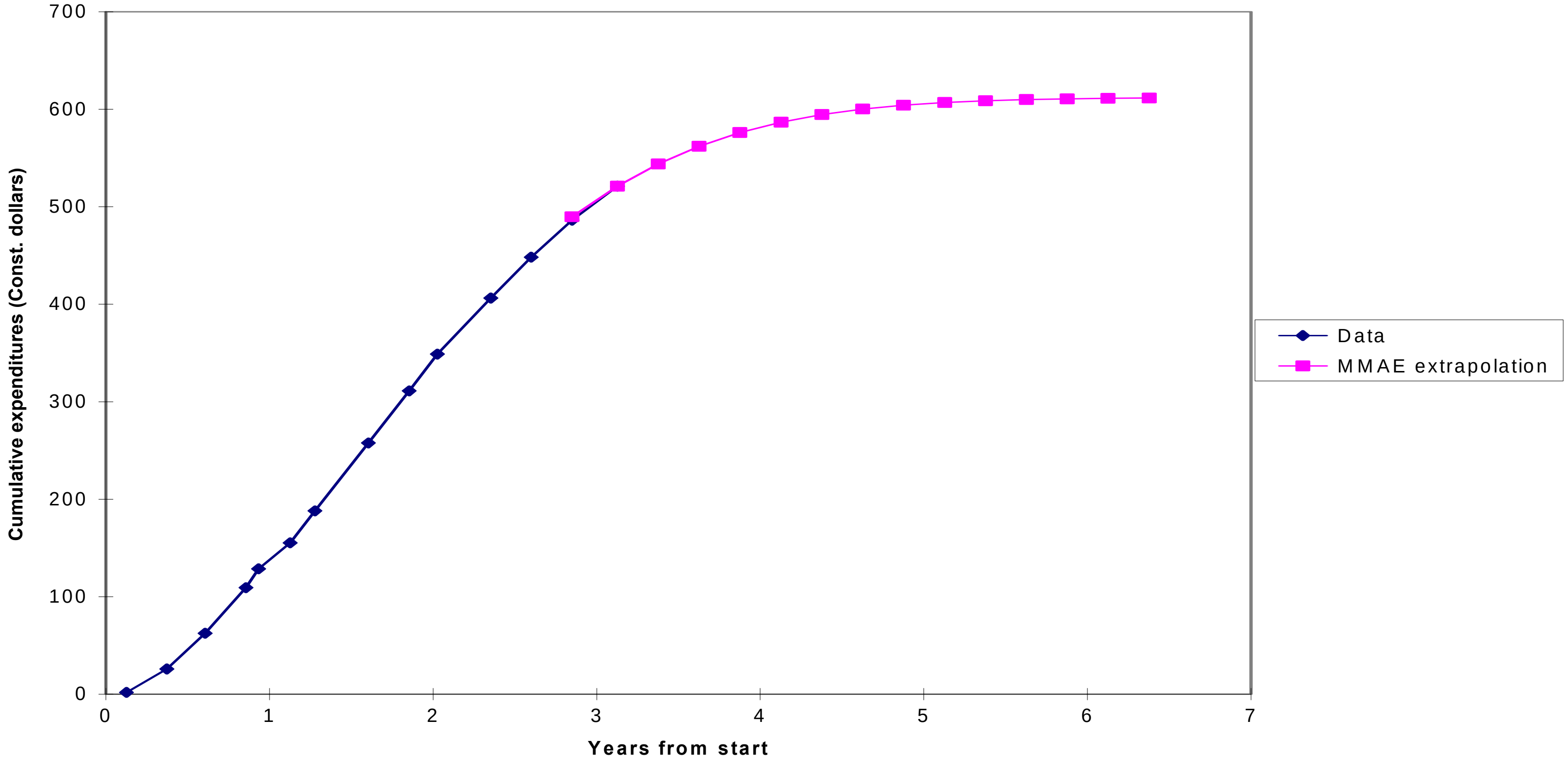
An Example

11/15/94	0
12/31/94	1.9
3/31/95	26.8
6/25/95	65.4
9/24/95	114.6
10/22/95	135.1
12/31/95	163.4
2/25/96	198.1
6/23/96	272.6
9/22/96	330
11/24/96	370.8
3/23/97	433.1
6/22/97	479
9/21/97	520.6
12/31/97	559

Optimal filter output and data

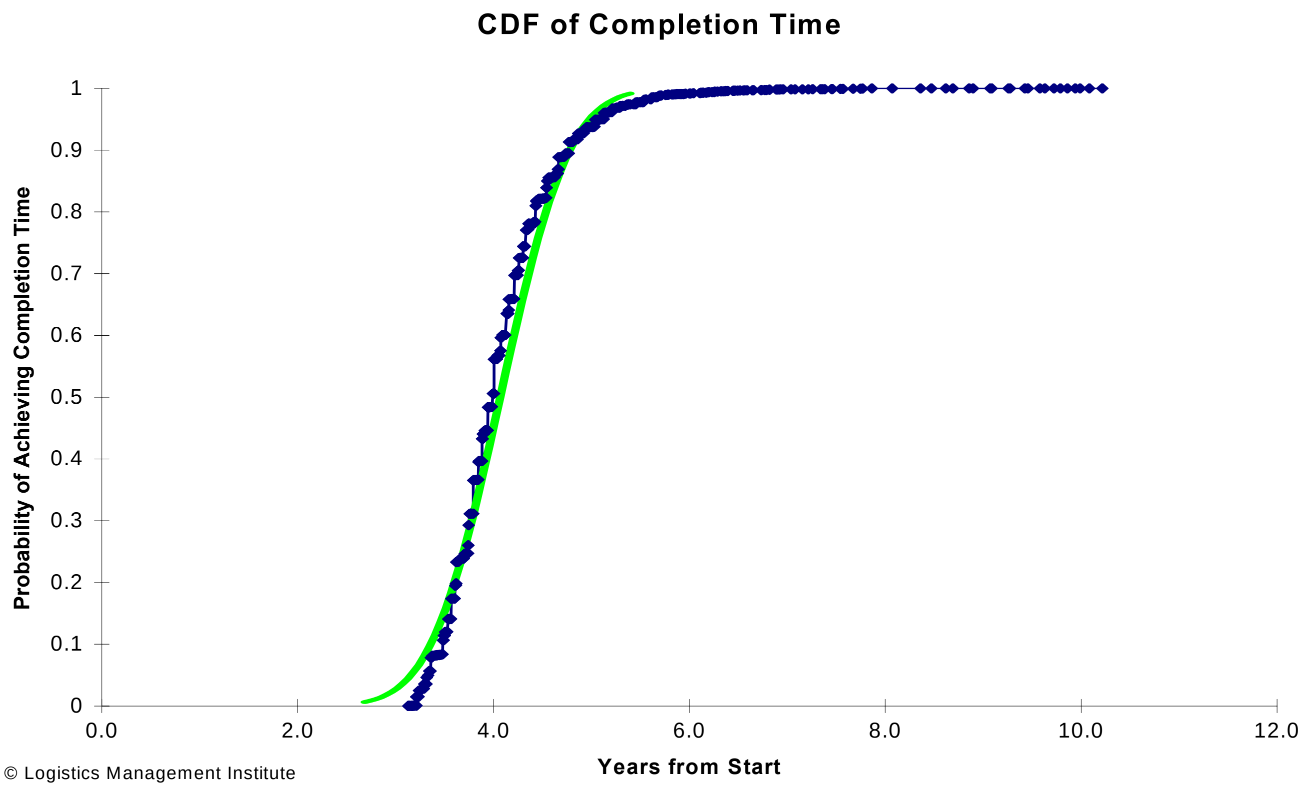


Norden-Rayleigh Extrapolation

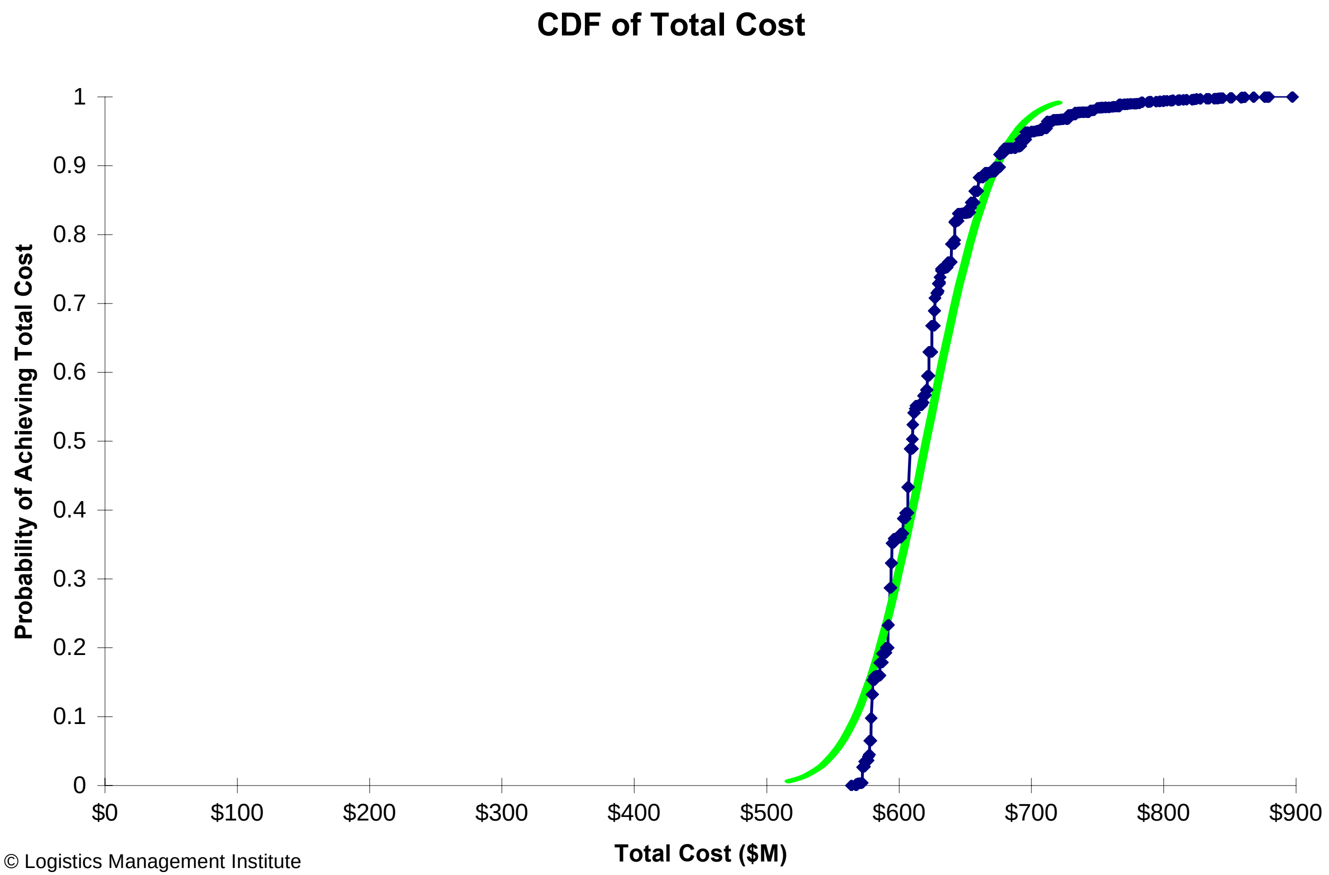


LMI
GOVERNMENT CONSULTING

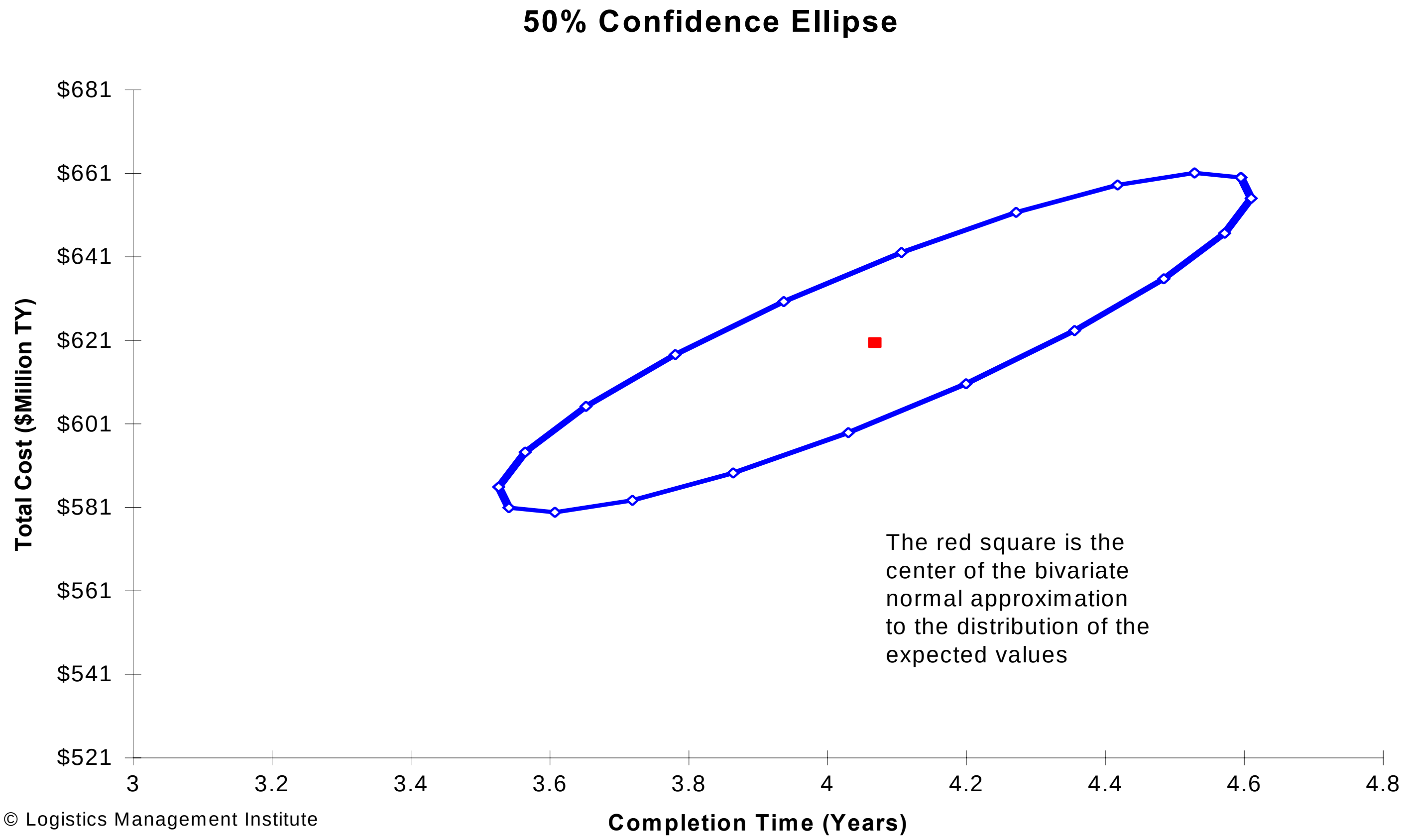
Marginal distribution of completion time



Marginal distribution of total cost



Bivariate distribution of cost and time



A General Point of View

- When making and using mathematical models for cost estimating and analysis, one should:
  - Base the models on economic or physical principles, OR
  - Base the models on clearly described empirical evidence
  - Develop and apply the models with careful mathematics and statistics
- Benefits:
  - Maximal useful output
  - Straightforward explanations of the work
  - Ability to use discrepancies to improve the models
- Cost estimating and analysis generally belongs to the discipline of system identification, and methods of this discipline are useful to cost analysts